



International Conference of Economics, Finance and Accounting Studies

International Conference of Economics, Finance and Accounting Studies is a double-blind peer-reviewed, open-access journal published to reach excellence on the scope. It considers scholarly, research-based articles on all aspects of economics, finance and accounting. As an international congress aimed at facilitating the global exchange of education theory, contributions from different educational systems and cultures are encouraged. It aims to provide a forum for all researchers, educators, educational policy-makers and planners to exchange invaluable ideas and resources.

Kredit To'lovini Bajarmaslik Xavfini Decision Tree Modeli Yordamida Baholash

Sattorov Abdumalik Abduraxmon o'g'li
O'zbekiston Respublikasi Markaziy banki
E-mail: abdumalikfinance@gmail.com

Annotatsiya: Tezisdagi kredit to'lovini bajarmaslik xavfini ma'lumotlar tahlili va Decision Tree modeli yordamida baholash usuli taklif etiladi. Maqsad to'lov tarixidan belgilar shakllantirish, model murakkabligini cheklash va tasniflash xatolarini bank vazifasiga mos baholash tartibini ishlab chiqishdir. Qarorlar daraxtining Gini mezon, precision, recall va F1 ko'rsatkichlari yoritiladi. Ilmiy-uslubiy hissa ma'lumotlar sizib chiqishining oldini olish, tushuntiriladigan qoidalar va xatolar qiymatiga mos qaror chegarasini yagona sinov tartibida bog'lashdan iborat. Ochiq ma'lumotlar bilan keyingi tajriba uchun reja hamda formulalarni tushuntiruvchi shartli hisob keltiriladi. Haqiqiy model o'qitilmagan va empirik samaradorlik natijalari olinmagan.

Kalit So'zlar: Data Science, Kredit Xavfi, Decision Tree, Gini Mezon, Precision, Recall, F1, Ma'lumotlar Sizib Chiqishi

Kirish

Kredit to'lovini bajarmaslik xavfini oldindan baholash bankka mijozlar bilan ishlash va tekshirish resurslarini rejalashtirishda yordam berishi mumkin. Data science yondashuvida to'lov tarixi, hisoblangan qarzdorlik va avvalgi to'lovlardan belgilar shakllantirilib, ular tasniflash modeliga uzatiladi. Qarorlar daraxti (Decision Tree) belgilar bo'yicha ketma-ket shartlar tuzadi. Daraxt yordamida kuzatuvlardan tasniflash qoidalarini hosil qilish yondashuvi Quinlan tadqiqotida yoritilgan [1].

Ushbu tezisning maqsadi kredit to'lovi xavfini Decision Tree orqali baholash uchun tekshiriladigan tajriba tartibini taklif etishdir. Modelning natijasi faqat umumiy to'g'ri tasniflash ulushi bilan emas, xavfli holatlarni aniqlash va noto'g'ri ogohlantirishlar o'rtasidagi nisbat bilan ham baholanadi. Ish nazariy-uslubiy xususiyatga ega; unda o'qitilgan modelning haqiqiy natijalari keltirilmaydi.

Ma'lumotlar va tadqiqot usuli

Keyingi metodik sinov uchun UCI Machine Learning Repository'dagi Default of Credit Card Clients to'plami taklif etiladi. Rasmiy tavsifga ko'ra, unda Tayvandagi mijozlarga oid 30 000 kuzatuv va 23 ta belgi mavjud; maqsadli o'zgaruvchi to'lov defaultini 1 va 0 bilan ifodalaydi [2].

Ushbu to'plam mahalliy bank ma'lumotlarini ifodalamaydi. Haqiqiy bank tajribasida prognoz davri va to'lovni bajarmaslik mezonini bank yozuvlariga mos ravishda oldindan aniq belgilanadi.[3] Boshlang'ich belgilar sifatida prognoz vaqtidan oldin mavjud bo'lgan moliyaviy ma'lumotlar olinishi taklif etiladi (1-jadval). Mijoz identifikatori model belgisi sifatida ishlatilmaydi. Kelgusi davrning to'lov natijasi faqat maqsadli o'zgaruvchi bo'ladi; uni yoki unga bevosita bog'liq kelajak yozuvini kirish belgilariga qo'shish taqiqlanadi.[4] Kategoriya kodlari ma'lumotlar lug'ati bilan tekshiriladi. Yetishmayotgan qiymatlarni to'ldirish va belgilarni tanlash qoidalari faqat o'qitish qismida aniqlanadi [5].

1-Jadval. Model uchun taklif etilayotgan belgilar guruhlariga[6]

Belgi guruhi	Ma'lumot	Modeldagi vazifasi
Kredit limiti	Mijozga berilgan limit miqdori.	Qarzdorlikni limitga nisbatan baholash uchun asos.
To'lov holati tarixi	Oldingi olti oydagi to'lov holati kodlari.	Avvalgi to'lov muammolari takrorlanishini ifodalash.
Hisoblangan qarzdorlik	Oldingi oylardagi hisob-kitob summalar.	Qarzdorlik hajmi va o'zgarishini ifodalash.
Amalda to'langan summa	Oldingi oylarda amalga oshirilgan to'lovlar.	To'lov xatti-harakatini ifodalash.

Manba: [2] dagi moliyaviy belgilar asosida muallif taklifi. Jadval empirik belgi muhimligi natijasi emas.

Bir mijozga bitta kuzatuvdan iborat to'plamda o'qitish, validatsiya va yakuniy test qismlarini 70:15:15 nisbatda, sinflar ulushini saqlab ajratish taklif etiladi. Bu nisbat uslubiy tanlov bo'lib, yagona majburiy qoida emas. Haqiqiy bankning oylik yozuvlarida vaqt bo'yicha ajratish qo'llanadi: oldingi davrlar o'qitish, keyingi davrlar tekshirish uchun olinadi. Bir mijozning takroriy yozuvlari va kelajak ma'lumotlari sababli paydo bo'ladigan bog'liqlik hisobga olinadi. Modelni Python muhitidagi scikit-learn vositalari bilan amalga oshirish mumkin [7]. Daraxt chuqurligi uchun 3, 5 va 7, bargdagi eng kam kuzatuvlar soni uchun 20, 50 va 100 boshlang'ich variantlari validatsiyada taqqoslanishi taklif etiladi. Sinflar nomutanosib bo'lsa, sinf vaznlari teng va o'qitish qismidagi chastotalarga mos vaznlash variantlari solishtiriladi [8]. Bu qiymatlar optimalligi isbotlangan sozlamalar emas. Yakuniy test model va qaror chegarasi tanlangandan keyin bir marta baholash uchun saqlanadi [9].

Decision Tree modelining matematik asosi

Gini mezoniga asoslangan ikkilik daraxtda tugundagi sinflar aralashligi quyidagicha ifodalanadi [10]:

$$G = 1 - p_0^2 - p_1^2 \quad (1)$$

Bu yerda p_0 va p_1 — tugundagi 0 va 1 sinflar ulushi; ularning yig'indisi 1 ga teng. Faqat bitta sinfdan iborat tugunda $G = 0$ bo'ladi. Tugunni ikkiga ajratish sifati vaznlangan kamayish orqali baholanadi:

$$\Delta G = G - \frac{N_L}{N} G_L - \frac{N_R}{N} G_R \quad (2)$$

Bu yerda G — ota tugundagi Gini qiymati; G_L va G_R — chap va o'ng tugundagi qiymatlar; N_L va N_R — ulardagi kuzatuvlar soni; $N = N_L + N_R$. Algoritm har tugunda ruxsat etilgan ajratishlar orasidan aralashlikni ko'proq kamaytiradigan variantni tanlaydi. Bu bosqichma-bosqich tanlov butun daraxtning global eng yaxshi yechimini kafolatlamaydi [11]. Sinf vaznlari qo'llansa, ulush

va sonlar o'rniga tegishli vaznlangan qiymatlar olinadi.

Daraxtning chuqurligi va barg hajmini cheklash murakkab qoidalarining o'qitish ma'lumotiga haddan tashqari moslashishini kamaytirishga qaratiladi [12]. Bargga olib boruvchi shartlar individual prognozning model ichidagi asosini ko'rsatadi. Ular sababiy bog'liqlikni isbotlamaydi. Bargdan olingan xavf skori alohida tekshirishsiz kalibrlangan default ehtimoli deb qabul qilinmasligi lozim.

Baholash mezonlari va shartli hisob

To'lovni bajarmaslik 1-sinf deb olinadi. TP — to'g'ri aniqlangan default holati; FN — aniqlanmay qolgan default; FP — default kuzatilmagan mijozga berilgan noto'g'ri ogohlantirish; TN — to'g'ri aniqlangan default bo'lmagan holat. Precision, recall va F1 quyidagicha hisoblanadi [13]:

$$\text{Precision} = \frac{TP}{TP + FP}; \text{Recall} = \frac{TP}{TP + FN}; F1 = \frac{2TP}{2TP + FP + FN} \quad (3)$$

Precision ogohlantirishlar orasidagi haqiqiy default ulushini, recall esa mavjud default holatlarining qanchasi aniqlanganini ko'rsatadi. F1 ularni birlashtiradi. Maxraj nol bo'lsa, ko'rsatkichni talqin qilish qoidasi oldindan belgilanadi. Yakuniy baholashda sinflar soni va to'liq xatolar matritsasi ham keltirilishi kerak.

2-Jadval. Formulalarni tushuntirish uchun shartli xatolar matritsasi

Haqiqiy holat	Prognoz: default (1)	Prognoz: default yo'q (0)
Default (1)	TP = 140	FN = 60
Default yo'q (0)	FP = 80	TN = 720

Manba: muallif tuzgan 1 000 kuzatuvli shartli misol. Sonlar UCI to'plamidan hisoblanmagan va o'qitilgan model natijasi emas.

Shartli misolda precision = $140/220 = 63,64\%$; recall = $140/200 = 70\%$; $F1 = 280/420 = 66,67\%$. Umumiy to'g'ri tasniflash ulushi (accuracy) = $860/1\,000 = 86\%$. Barcha holatlarni «default yo'q» deb belgilaydigan oddiy qoida shu misolda 80% accuracy beradi, ammo defaultni aniqlash darajasi 0% bo'ladi. Bu hisob accuracy'ni boshqa ko'rsatkichlar bilan birga talqin qilish zaruratini tushuntiradi.

Bank vazifasida FN va FP oqibatlari farqlanishi mumkin. Qaror chegarasini tanlash uchun validatsiya qismida quyidagi shartli jarima mezonni qo'llanishi taklif etiladi:

$$C = c_{FN} \times FN + c_{FP} \times FP \quad (4)$$

Bu yerda c_{FN} va c_{FP} — bank belgilaydigan xato qiymatlari, C esa umumiy jarima mezonidir. Masalan, $c_{FN} = 5$ va $c_{FP} = 1$ shartli birlik bo'lsa, 2-jadval uchun $C = 5 \times 60 + 1 \times 80 = 380$ birlik. Bu haqiqiy pul zarari hisob-kitobi emas. Xavf skori asosida 1-sinfga o'tkazish chegarasi validatsiyada ushbu mezonga mos tanlanadi; test to'plami chegara tanlashga jalb etilmaydi [14].

Ilmiy-uslubiy hissa va amaliy taklif

Tezisning ilmiy-uslubiy hissasi kredit to'lovi tarixidan belgilar shakllantirish, kelajak ma'lumotining o'qitishga sizib chiqishini cheklash, daraxt murakkabligini boshqarish va xatolar qiymatiga mos qaror chegarasini bir sinov tartibida bog'lashdan iborat. Bu hissa bank vazifasiga moslashtirilgan uslubiy sintezdir; Decision Tree algoritmi yoki baholash formulalari yangi algoritm sifatida taklif etilmaydi.

Amaliy sinovda model barcha holatlarni eng ko'p uchraydigan sinfga ajratuvchi sodda boshlang'ich qoida bilan bir xil testda taqqoslanadi. Natijalar precision, recall, F1 va xatolar matritsasi bilan beriladi; daraxt chuqurligi va barglar hajmi ham qayd etiladi. Yuqori xavf skori

kreditni avtomatik rad etish uchun tayyor asos sifatida olinmaydi: u xodim tekshiruv va mijoz bilan ishlash ustuvorligini belgilashga yordam beruvchi signal bo'lishi mumkin.[15]

Xulosa

Decision Tree kredit to'lovi xavfini jadval ma'lumotlari asosida baholash va prognozni shartlar orqali izohlash uchun mos boshlang'ich modeldir. Tezisda ma'lumotlarni ajratish, Gini mezon orqali daraxt qurish, model murakkabligini cheklash va xatolar oqibatiga mos baholash tartibi taklif etildi. Shartli hisob umumiy accuracy yuqori bo'lsa ham default holatlarini aniqlash alohida tekshirilishi kerakligini ko'rsatadi. Keyingi vazifa ushbu tartibni haqiqiy ma'lumotlarda sinash, prognoz barqarorligini baholash va mahalliy bank sharoitiga moslashtirishdir.

Foydalanilgan Adabiyotlar

- [1] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, 1986, doi: 10.1007/BF00116251.
- [2] I.-C. Yeh, "Default of Credit Card Clients," *UCI Machine Learning Repository*, 2009, doi: 10.24432/C55S3H.
- [3] scikit-learn, "Decision Trees," *scikit-learn Documentation*, [Online]. Available: scikit-learn official documentation. Accessed: Oct. 6, 2026.
- [4] scikit-learn, "DecisionTreeClassifier," *scikit-learn Documentation*, [Online]. Available: scikit-learn official documentation. Accessed: Oct. 6, 2026.
- [5] scikit-learn, "Common pitfalls and recommended practices," *scikit-learn Documentation*, [Online]. Available: scikit-learn official documentation. Accessed: Oct. 6, 2026.
- [6] scikit-learn, "Metrics and scoring: Quantifying the quality of predictions," *scikit-learn Documentation*, [Online]. Available: scikit-learn official documentation. Accessed: Oct. 6, 2026.
- [7] scikit-learn, "Tuning the decision threshold for class prediction," *scikit-learn Documentation*, [Online]. Available: scikit-learn official documentation. Accessed: Oct. 6, 2026.
- [8] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.
- [9] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann, 1993, 302 p.
- [10] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [11] D. J. Hand and R. J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems," *Machine Learning*, vol. 45, no. 2, pp. 171–186, 2001, doi: 10.1023/A:1010920819831.
- [12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.
- [13] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, no. 3, Art. no. e0118432, 2015, doi: 10.1371/journal.pone.0118432.
- [14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009, 745 p., doi: 10.1007/978-0-387-84858-7.
- [15] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, pp. 4765–4774, 2017.